I J C T A, 9(14) 2016, pp. 6809-6814

Brain Wave Analysis Towards Familiar and Unfamiliar Language

D. Sasi Rekha*

ABSTRACT

Many of the individuals we know, are familiar with more than one language. However, the degree of comfort varies. Hence theresearch question is whether the brain activity is less during its encounter with more familiar language. With the motivation tofind an answer, the current study aims to statistically interpret the response behavior of two individuals towards Tamil andEnglish. The raw data of brainwaves collected using Electroencephalography (EEG reader) was transformed and analyzed forthe significant differences, in the response behaviors of individuals towards respective languages.

INTRODUCTION

Brainwave is dependent on electrical activity in the brain and is in-turn dependent on an individual's activity. Brainwavescontain a spectrum of frequencies response towards its associated activities. The frequencies can be as summarized in Table 1.

Table 1
Brainwave frequencies and associated activities

Brainwave Frequency	**Activities**	
Gamma	*27 Hz and above*	*Formation of ideas & language processing*
Beta	12 Hz – 27 Hz	Alert
Alpha	8 – 12 Hz	Awake and relaxed
Theta	3 – 8 Hz	Extreme relaxed
Delta	0.2 – 3 Hz	Deep and dreamless sleep

The contribution of EEG techniques in understanding how the brain processes information from language is huge. One methodto understand and gain information is via coherence computation. (Weiss & Mueller, 2003) The method can be used to measurethe linear dependency between two distant brain regions. There has been research into brainwave entrapment to improve tasklearning and memory retaining. (Will & Berg, 2007) One of the aspects in task learning is learning new languages.

Reiterer *et al.* (2005) did a study on English and German speakers of different proficiency. Significant differences were detectedin the network activity between high and low proficiency groups (S. Reiterer, Hemmelmann, Rappelsberger, & Berger, 2005).In 2009, Reiterer et. al. did a similar study and found out that the right brain was significantly more involved in second languagein our less proficient second language speakers. However, this observation was indicated in long-range synchronization patternsin the gamma frequency range.(Susanne Reiterer, Pereda, & Bhattacharya, 2009).

* Assistant Professor, Department of Electrical and Electronics Engineering, SRM University, Kattankulathur, India, *E-mail: rekhasasi123@gmail.com*

Few research efforts look into the EEG study in English and Tamil, which is similar to the current study's focus. Studies have shown that reading English and Tamil seem to have commonresponse mechanisms. However, further analysis show that there is a difference in extra brain areas, which are usually not oftenobserved in alphabetic studies (Liu & Perfetti, 2003).

The purpose of this study is to check for significant differences between brainwaves responses in two speakers while readingarticles in English and Tamil. Thus there are two hypotheses:

(1) Reading a more familiar language is easier and hence produces less brain response when compared with the brain'sresponse while reading less familiar language.

(2) Different people respond with different brainwave intensities when reading the same language.

20 readings with 2 minutes each is recorded using the EEG. A variety of statistical tools are used to check for the presencesignificant differences. For this study, Igor pro software has been used to perform various statistical analysis. Following sectionsdescribe the procedure adopted during data collection and its analysis.

Statistical Theories used: The non-normality of data are checked graphically by plotting a Quantile-Quantile (QQ plot) or thenormality probability plot. The data is said to be normal if the plot is a straight line. If the data were normal, then Student Ttestis the best way to get an estimate of the sample mean. However, if the data is non-normal then some of the methods thatcan be followed are as mentioned below:

- Lognormal distribution: If the Logarithm of the data in hand follows a normal distribution then data is called as lognormal. Also, depending on the confidence of this normality, further there are different ways to estimate.
- Bootstrap t-interval for the mean: This is carried out by resampling from the data as if it were the population. The resultingdata is again checked for normality.

Comparison of means of two samples: T-test is normally used for comparing two sample means having normal distribution. Inwhich case, if the P-value estimate is less than alpha, (% significance ~95% for this project i.e., $\alpha = 0.05$ for all comparison)then there is said to be significant difference.

In case of correlation between the two compared samples (samples are dependent), then paired t-test is carried out to nullifythe influence of any of the common factors present. However, for comparing non normal sample means, Wilcoxon Rank Sum Test is an alternative. This test assumes that the twodistributions are identical under null hypothesis and identical, except for a shift under alternative hypothesis. In this case thetest can be interpreted as a test of equal means or test of equal medians. Also, in case of dependent samples, Wilcoxon SignedRank test is non-parametric alternative to the paired t-test. While carrying out this test, the difference between the two data areordered by their absolute value and assigned ranks to the differences. These ranks are then summed by the sign of differencesto give Tp (sum of positive rank) and Tn (sum of negative rank). The null hypothesis tested by this procedure is whether thedistribution of the differences is symmetric about zero or a particular value. Hence if the smaller of the Tp or Tn is less thancritical value at alpha (α=0.05), then the null hypothesis is rejected (STAT 511, Experimental Design and Data Analysis forResearchers I, Colorado State University, Fall 2009, Instructor: Dr. Daniel Cooley).

METHODOLOGY

The Null hypotheses of this study are:

- There is no significant difference between brain responses to different languages by an individual operator
- There is no significant difference between brain responses of different operator towards a particular language.

To test the hypotheses, four sets of data were collected by EEG from two subjects, Subject C and Subject M. The four sets ofdata listed below, contains 20 brainwaves, recorded for 120 seconds:

I. C is reading Tamil.

II. M is reading Tamil.

III. C is reading English.

IV. M is reading English.

The applied procedure of this project is as summarized in Figure 1:

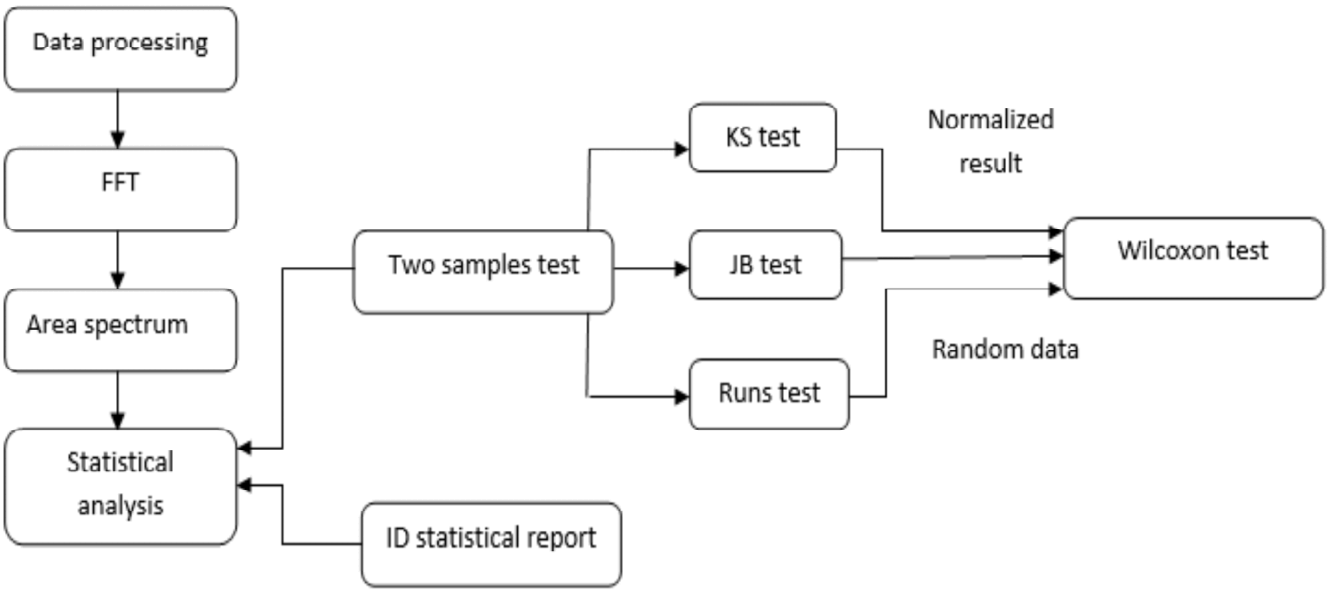

Figure 1: Technical routes for EEG brainwave analysis

(1) Pre-processing of the raw data. After the collection of the raw data, it was found that the data length over the same periodcontained different numbers of points. One of the possible reasons could be the presence of a lag period before the actualconnection establishment between EEG and the computer. This lag period is estimated to be different for every data3collection cycle of 120 seconds. In order to keep the constant time step, every brainwave data was matched with theshortest length brainwave. This data length was matched by deletion of some data at the beginning part of every brainwave,which is believed to be a part of the delay.

(2) Fast Fourier Transform (FFT) of the brainwaves. EEG collects data in time domain.But spectrums of brainwavesdisplay more information about brain activities, so we need to transform the data into frequency domain before analysis.By doing FFT, we got 80 spectrums of the 4 groups mentioned earlier.

(3) Area under the spectrum curve. In this project, spectrum areas indicate the brain's response intensities. The spectrumareas above 27 Hz were calculated, as the hypothesis of interest is to judge the difference between the response behaviorstowards different languages (Table 1). Therefore, 20 spectrum areas for every condition were obtained.

(4) Statistical analysis. The data of the individual group were analyzed using an in-built feature of the software Igor Pro,called 1D Statistics report. This was carried out, in order to understand the data distributions. Two samplecomparison was carried out using "Two sample test" module in the software. (Consolidated report in Table 2, 3 and in Table 4, 5).

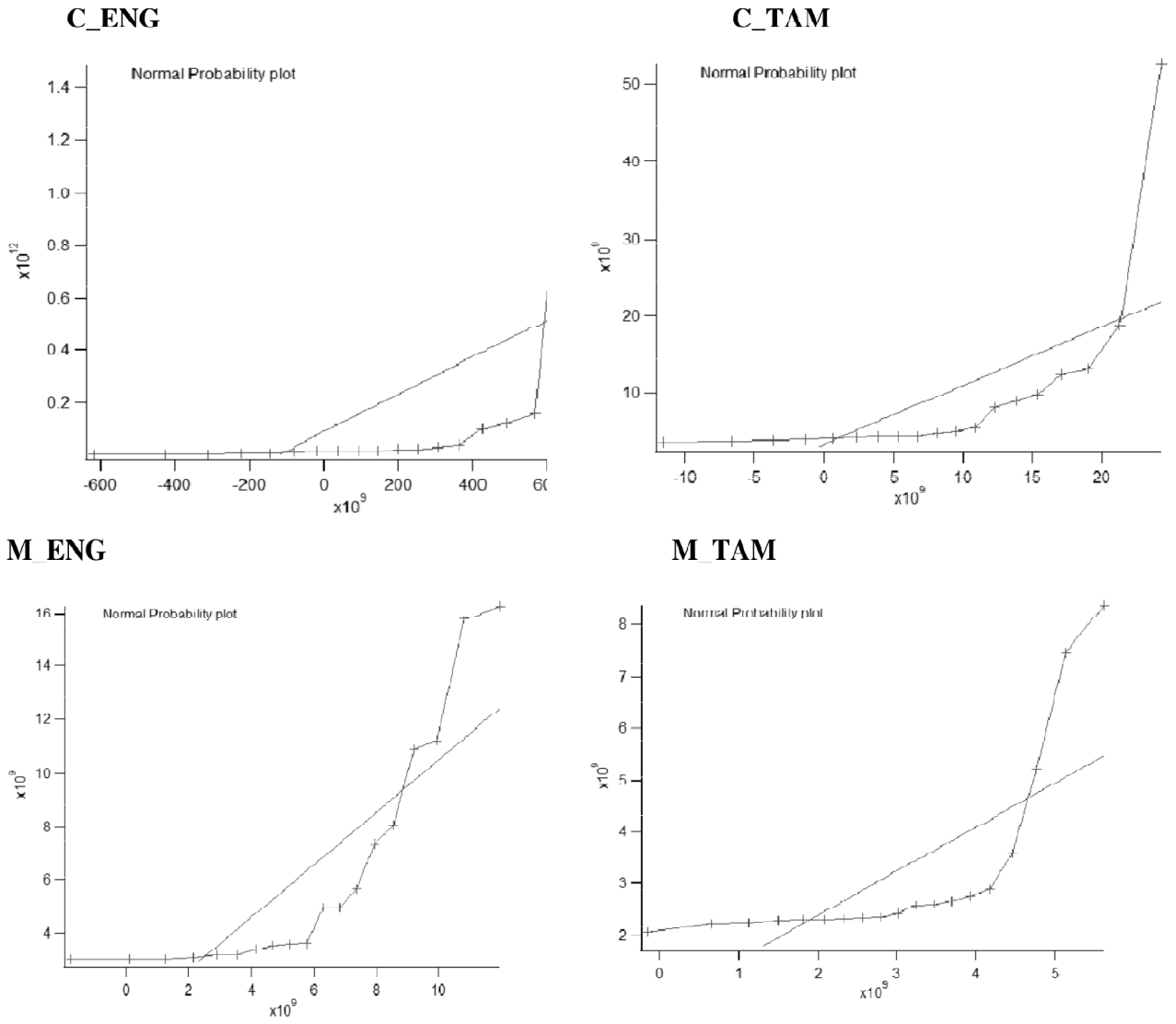

Figure 2: QQ plot-Normality Plot of respective group

RESULTS & DISCUSSION

Figure 2 displays the QQ-plot or the normality plot for the four groups. As it can be seen, neither of the sample groups have astraight line. Thereby it can be said that the data is not normal. The non-normal sample means are compared using WilcoxonSigned Rank test.

Table 2 compares the difference between a subject's response to English and Tamil. The respective p-values (two tail) of Csubject and M subject are 0.001 and 0.0002 which are less than α=0.05. Hence implying that the subjects' response todifferent languages is significantly different. P_lower tail checks for the hypothesis that Tamil response is lower thanEnglish response for a given particular subject. With the P_lower tail lower than α, it can be rejected that Tamil response islower than English response for both subjects.

Likewise, Table 3 compares the subjects' response to the same language, i.e., comparison between C & M subject for Tamil and English respectively (P_values (2 tail) <0.001). Hence it can be said that each subject responds differentlytowards a particular language. P_lower tail checks for the hypothesis that C's response is lower than M's response for Tamil and English respectively. With the P_lower tail higher than α, it can be accepted that subject C's response is lowerthan M's response. The P_Upper tail tests for the opposite hypothesis as the P_lower tail.

Table 2
Tested Hypothesis: No difference between thelanguages for a given subject

Wilcoxon Signed Rank Paired T – test of Tamil Vs English		
Parameters	*C_ Subject*	*M_Subject*
N	20	20
Sum of Positive rank, Tp	22	195
Sum of negative rank, Tn	188	15
P_Lower tail	0.0005	0.00013
P_Upper tail	0.99	0.99
P_Two_ tail	0.001	0.00026

Table 3
Tested Hypothesis: No difference between theSubjects for a given Language

Wilcoxon Rank Sum test of C Vs M Subject [Independent]		
Parameters	*Tamil*	*English*
m	20	20
n	20	20
Total points	40	40
U Statistic	359	352
Up Statistic	41	48
P_Lower tail	0.99	0.99
P_Upper tail	0.00000193	0.00000594
P_Two_ tail	0.00000386	0.0000119
W		

CONCLUSION

This project's focus of interest was to study the response behavior of human subjects towards different languages. It can bestatistically concluded that irrespective of the familiarity with languages, an individual brainwave response is different.

Statistically, the Tamil response is higher than English response for both subjects. In other words, both subject are morecomfortable with English. This is because, the response intensity is expected to be lower for easier or more comfortableactivities(S. Reiterer et al., 2005)Statistically, the subject C's response is lower than M's response for both languages. In other words, Subject C is morecomfortable with the languages than subject M.

In perceived reality, M is more proficient in Tamil than in English and hence is expected to be more comfortable with Tamil.But the study shows that M is more comfortable with English. One of the possible reasons could be the significant instrumentvariation while collecting data. Also, in real time, C is more proficient in English as against thestatistical information. Overall, it can be concluded that the general expectation of less response towards familiar languagecannot be corroborated statistically.

REFERENCES

[1] Chinese University of Hongkong. Research on Language and Brain Waves. from *http://www.ee.cuhk.edu.hk/~lel/shiae/Research.html*

[2] Liu, Y., & Perfetti, C. A. (2003). The Time Course of Brain Activity in Reading English and Chinese: An ERP Study ofChinese Bilinguals. *Human Brain Mapping, 18*, 167–175.

[3] Reiterer, S., Hemmelmann, C., Rappelsberger, P., & Berger, M. L. (2005). Characteristic functional networks in high-vs.low-proficiency second language speakers detected also during native language processing: an explorative EEGcoherence study in 6 frequency bands. *Cognitive Brain Research, 25*, 566–578.

[4] Reiterer, S., Pereda, E., & Bhattacharya, J. (2009). Measuring second language proficiency with EEG synchronization: howfunctional cortical networks and hemispheric involvement differ as a function of proficiency level in secondlanguage speakers. *Second Language Research, 1*, 77-106.

[5] Weiss, S., & Mueller, H. M. (2003). The contribution of EEG coherence to the investigation of language. *Brain andLanguage, 85*, 325–343.

[6] Will, U., & Berg, E. (2007). Brain wave synchronization and entrainment to periodic acoustic stimuli. *Neuroscience Letters,424*, 55–60.

[7] STAT 511, Experimental Design and Data Analysis for Researchers I, Colorado State University, Fall 2009, Instructor: Dr.Daniel Cooley.